On the Distribution of Products of Primes and Powers

Adrian W. Dudek

Mathematical Sciences Institute

The Australian National University

adrian.dudek@anu.edu.au

Abstract

We prove several results regarding the distribution of numbers that are the product of a prime and a k-th power. First, we prove an asymptotic formula for the counting function of such numbers; this generalises a result of E. Cohen. We then show that the error term in this formula can be sharpened on the assumption of the Riemann hypothesis. Finally, we prove an asymptotic formula for these counting functions in short intervals.

1   Introduction

It is the purpose of this paper to understand further the distribution of numbers of the type pm^k, where p is a prime and m and k are positive integers. The case $k = 2$ was considered by Cohen [1] in 1962, and it was therein established that

$$\sum_{pm^2 \le x} 1 = \zeta(2)\frac{x}{\log x} + O\left(\frac{x}{\log^2 x}\right)$$

where the sum is over primes p and positive integers $m \ge 1$. As it is known that every number can be uniquely represented as the product of a square-free number and a square, it follows that the sum in the above equation counts each number pm^2 precisely once. Moreover, every number has a unique representation as the product of a k-free number and a k-th power, and so we

1

define
$$C_k(x) := \sum_{pm^k \leq x} 1.$$

Throughout this paper we call on standard results from analytic number theory. We will refer the reader to the appropriate section of Montgomery and Vaughan [2] for further details.

Our first objective is to furnish the following theorem.

Theorem 1. *Let $k \geq 2$ and let $C_k(x)$ count those numbers not exceeding x that can be represented as the product of a prime and a k-th power. Then there exists a constant $c > 0$ such that*

$$C_k(x) = \zeta(k)li(x) + O\left(x \exp\left(-c\frac{(\log x)^{3/5}}{(\log\log x)^{1/5}} \right) \right).$$

Considering that the set of numbers of the form pm^k includes the prime numbers, it is apparent that any improvement to the error term in Theorem 1 will require a sharper form of the zero-free region. One can also reduce the error term in the above theorem by assuming the Riemann hypothesis to be true; this is the assertion that all non-trivial zeroes ρ of the Riemann zeta-function satisfy $\text{Re}(\rho) = 1/2$.

Theorem 2. *Assume the Riemann hypothesis. Let $k \geq 2$ and let $C_k(x)$ count those numbers not exceeding x that can be represented as the product of a prime and a k-th power. Then we have that*

$$C_2(x) = \zeta(2)li(x) + O(x^{1/2}\log^2 x)$$

and

$$C_k(x) = \zeta(k)li(x) + O(x^{1/2}\log x)$$

for all $k \geq 3$.

Remark 1. *It seems curious that the case $k = 2$ should evade the error term which appears in the general case of Theorem 2. It would be interesting to see if somebody could prove the bound*

$$C_2(x) = \zeta(2)li(x) + O(x^{1/2}\log x)$$

on the Riemann hypothesis.

From Theorem 2, one can count numbers of the type pm^k in the interval $(x, x+h)$, where $h = o(x)$. Clearly, we have that

$$\sum_{x < pm^2 \leq x+h} 1 = C_2(x+h) - C_2(x) = \zeta(2) \int_x^{x+h} \frac{dt}{\log t} + O(x^{1/2} \log^{A(k)} x)$$

where $A(2) = 2$ and $A(k) = 1$ for $k \geq 3$. From the estimate

$$\int_x^{x+h} \frac{dt}{\log t} \gg \frac{h}{\log(x+h)},$$

we have that the asymptotic formula

$$C_k(x+h) - C_2(x) \sim \zeta(k) \int_x^{x+h} \frac{dt}{\log t}$$

holds for all $k \geq 2$ provided that $h(x)/(x^{1/2} \log^{A(k)+1} x) \to \infty$.

Using an explicit formula that relates the distribution of primes to the zeroes of the Riemann zeta-function, we can improve these estimates as demonstrated in the following theorem.

Theorem 3. *Assume the Riemann hypothesis. Then we have that the asymptotic formula*

$$C_k(x+h) - C_k(x) \sim \zeta(k) \int_x^{x+h} \frac{dt}{\log t}$$

holds for all $k \geq 2$ provided that $h(x)/(x^{1/2} \log^{A(k)} x) \to \infty$, where $A(2) = 2$ and $A(k) = 1$ for all $k \geq 3$.

It is apparent that any improvement to Theorems 2 and 3 for the case where $k \geq 3$ will require more information on the ordinates of the zeroes of the Riemann zeta-function. That is, these are best possible on the Riemann hypothesis. The author is unsure, however, as to whether one could improve the conditional estimates for $k = 2$, and so this can be considered as an open problem.

2 Proofs

2.1 Proof of Theorem 1

Let $k \geq 2$. It is convenient to work with the von Mangoldt function

$$\Lambda(n) = \begin{cases} \log p & : \quad n = p^m, \ p \text{ is prime}, \ m \in \mathbb{N} \\ 0 & : \quad \text{otherwise} \end{cases}$$

3

and consider the weighted sum

$$C_k^*(x) := \sum_{nm^k \leq x} \Lambda(n).$$

It should be remarked that $C_k^*(x)$ will count some numbers (such as 36) more than once. We will attenuate this contribution later on.

Lemma 4. *Let $k \geq 2$ and let*

$$C_k^*(x) = \sum_{nm^k \leq x} \Lambda(n).$$

Then there exists a constant $c > 0$ such that

$$C_k^*(x) = \zeta(k)x + O\left(x \exp\left(-c\frac{(\log x)^{3/5}}{(\log\log x)^{1/5}} \right) \right).$$

Proof. We divide this into two sums

$$C_k^*(x) = \sum_{\substack{nm^k \leq x \\ m \leq x^{1/2k}}} \Lambda(n) + \sum_{\substack{nm^k \leq x \\ m > x^{1/2k}}} \Lambda(n). \tag{1}$$

It is known (see Equation (6.28) of [2] for example) that

$$\sum_{n \leq x} \Lambda(n) = x + O\left(xe^{-\delta(x)} \right)$$

where

$$\delta(x) := c\frac{(\log x)^{3/5}}{(\log\log x)^{-1/5}} \tag{2}$$

for some constant $c > 0$. Therefore, one may estimate the first sum by

$$
\begin{aligned}
\sum_{\substack{nm^k \leq x \\ m \leq x^{1/2k}}} \Lambda(n) \;&=\; \sum_{m \leq x^{1/2k}} \sum_{n \leq x/m^k} \Lambda(n) \\
&=\; \sum_{m \leq x^{1/2k}} \left(\frac{x}{m^k} + O\left(\frac{x}{m^k} e^{-\delta(x/m^k)} \right) \right) \\
&=\; \zeta(k)x + O\left(x \sum_{m > x^{1/2k}} \frac{1}{m^k} \right) + O\left(x \sum_{m \leq x^{1/2k}} \frac{e^{-\delta(x/m^k)}}{m^k} \right).
\end{aligned}
$$

4

Clearly, the sum in the first error term can be bounded by comparison to the integral *viz.*

$$\sum_{m>x^{1/2k}} \frac{1}{m^k} \ll \int_{x^{1/2k}}^{\infty} t^{-k} \ll x^{-1/2+1/2k}.$$

For the second error term, we use that fact that $m \le x^{1/2k}$ to get that

$$\sum_{m\le x^{1/2k}} \frac{e^{-\delta(x/m^k)}}{m^k} \ll e^{-\delta(x^{1/2})} \sum_{m\le x^{1/2k}} \frac{1}{m^k} \ll e^{-\delta(x^{1/2})}.$$

Therefore, we have that

$$\sum_{\substack{nm^k\le x \\ m\le x^{1/2k}}} \Lambda(n) = \zeta(k)x + O(e^{-\delta(x^{1/2})}).$$

It now remains to bound the second sum in (1). As $m > x^{1/2k}$, it follows that the sum will only be over the prime powers n with $n \le x^{1/2}$. Therefore, we have that

$$\sum_{\substack{nm^k\le x \\ m>x^{1/2k}}} \Lambda(n) \ll \sum_{n\le x^{1/2}} \Lambda(n) \sum_{m^k\le x/n} 1 \ll x^{1/2} \sum_{n\le\sqrt{x}} \frac{\Lambda(n)}{n^{1/k}}.$$

It follows by partial summation and the prime number theorem that

$$\sum_{\substack{nm^k\le x \\ m>x^{1/2k}}} \Lambda(n) \ll x^{1/2+1/2k}.$$

This completes the proof of the lemma. ☐

Remark 2. *One should note that the value of c in Lemma 4 will not necessarily be the same as that in (2). This change arises from the fact that $\delta(x^{1/2}) \ll \delta(x)$.*

We now prove Theorem 1 directly. Clearly, we may write

$$\sum_{pm^k\le x} \log p \;=\; \sum_{nm^k\le x} \Lambda(n) - \sum_{\substack{p^r m^k\le x \\ r\ge 2}} \log p.$$

By Lemma 4, it follows immediately that

$$\sum_{pm^k\le x} \log p \;=\; \zeta(k)x + O\!\left(xe^{-\delta(x)}\right) + O\!\left(\sum_{\substack{p^r m^k\le x \\ r\ge 2}} \log p\right)$$

5

where $\delta(x)$ is as before for some $c > 0$. We need to estimate the rightmost sum in the above equation. Clearly, we have

$$\sum_{\substack{p^r m^k \leq x \\ r \geq 2}} \log p \ll \sum_{m \leq x^{1/k}} \sum_{r \geq 2} \sum_{p \leq (x/m^k)^{1/r}} \log p.$$

By the prime number theorem, we have

$$\sum_{\substack{p^r m^k \leq x \\ r \geq 2}} \log p \ll \sum_{m \leq x^{1/k}} \sum_{r \geq 2} \left(\frac{x}{m^k} \right)^{1/r}$$

$$\ll x^{1/2} \sum_{m \leq x^{1/k}} \frac{1}{m^{k/2}}.$$

Clearly, the sum in the above formula is $O(\log x)$ when $k = 2$, and $O(1)$ for $k \geq 3$. Therefore, we have that

$$\sum_{pm^k \leq x} \log p = \zeta(k)x + O\left(xe^{-\delta(x)} \right)$$

It follows by partial summation that

$$\sum_{pm^k \leq x} 1 = \zeta(k)\mathrm{li}(x) + O(xe^{-\delta(x)}).$$

2.2 Proof of Theorem 2

Assume the Riemann hypothesis. It follows (see Theorem 13.1 of [2]) that

$$\psi(x) = x + O(x^{1/2} \log^2 x).$$

Working similarly to before, we have that

$$C_k^*(x) = \sum_{nm^k \leq x} \Lambda(n)$$

$$= \sum_{m \leq x^{1/k}} \psi\left(\frac{x}{m^k} \right)$$

$$= x \sum_{m \leq x^{1/k}} \frac{1}{m^k} + O\left(x^{1/2} \log^2(x) \sum_{m \leq x^{1/k}} \frac{1}{m^{k/2}} \right).$$

6

We can deal with the first sum as before to get

$$\sum_{m \le x^{1/k}} \frac{1}{m^k} = \zeta(k) + O(x^{1/k-1}).$$

Thus, we have that

$$C_k^*(x) = \zeta(k)x + O\left(x^{1/2} \log^2(x) \sum_{m \le x^{1/k}} \frac{1}{m^{k/2}}\right).$$

If $k = 2$, then the sum in the above equation is $O(\log x)$; otherwise, it is $O(1)$. To complete the proof, one simply needs to remove the contribution of powers of primes from $C_k^*(x)$ and apply partial summation as in the proof of Theorem 1.

2.3   Proof of Theorem 3

We define the weighted sum

$$\psi_1(x) := \sum_{n \le x} (x - n)\Lambda(n) = \int_2^x \psi(t)dt$$

and consider the well-known (see Equation (13.7) of [2]) explicit formula

$$\psi_1(x) = \frac{x^2}{2} - \sum_\rho \frac{x^{\rho+1}}{\rho(\rho+1)} - \frac{\zeta'}{\zeta}(0)x + \frac{\zeta'}{\zeta}(-1)x + O(x^{-1}) \qquad (3)$$

where the sum is over the non-trivial zeroes $\rho = \beta + i\gamma$ of the Riemann zeta-function $\zeta(s)$. Suppose that $2 \le \Delta \le h \le x$. We define a weight function $w_{x,h,\Delta}(n)$ viz.

$$w_{x,h,\Delta}(n) = \begin{cases} (n - x + \Delta)/\Delta & : & x - \Delta \le n \le x \\ 1 & : & x \le n \le x + h \\ (x + h + \Delta - n)/\Delta & : & x + h \le n \le x + h + \Delta \\ 0 & : & \text{otherwise.} \end{cases}$$

This function assumes the shape of an isosceles trapezoid, supported on the interval $(x - \Delta, x + h + \Delta)$, and constantly equal to 1 on the interval $(x, x+h)$. One can use such a weight to study the distribution of primes in short intervals, with better error terms than a weight with a sharp cut-off. As such, we define the sum

$$S_\Delta(x, h) = \sum_n \Lambda(n) w_{x,h,\Delta}(n).$$

7

The following lemma connects this sum with the distribution of zeroes of the Riemann zeta-function.

Lemma 5. *Let $2 \le \Delta \le h \le x$. Then*

$$S_\Delta(x, h) = h + \Delta - \frac{1}{\Delta} \sum_\rho S(\rho) + O\left(\frac{1}{\Delta x}\right)$$

where

$$S(\rho) := \frac{(x + h + \Delta)^{\rho+1} - (x + h)^{\rho+1} - x^{\rho+1} + (x - \Delta)^{\rho+1}}{\rho(\rho + 1)}.$$

Proof. One can confirm the identity

$$\sum_n \Lambda(n) w_{x,h,\Delta}(n) = \frac{1}{\Delta}(\psi_1(x + h + \Delta) - 2\psi_1(x + h) - \psi_1(x) - \psi_1(x - \Delta)).$$

by expanding the sums on the left hand side. From here, it remains to apply the explicit formula (3). $\qquad\square$

Let $m \ge 1$ and $k \ge 2$. The sum $S_{\Delta/m^k}(x/m^k, h/m^k)$ counts the number of prime powers (with weight) in the interval $((x - \Delta)/m^k, (x + h + \Delta)/m^k)$. This is equal to the count of numbers $p^r m^k$ in the interval $(x - \Delta, x + h + \Delta)$, though it is a small matter to later remove the contribution from higher prime powers as well as numbers outside of the interval $(x, x + h)$. Therefore, we are interested in the sum

$$\sum_{x-\Delta < nm^k < x+h+\Delta} w_{x,h,\Delta}(n)\Lambda(n) = \sum_{m^k < h} S_{\Delta/m^k}(x/m^k, h/m^k).$$

A direct application of Lemma 5 gives

$$\sum_{x-\Delta < nm^k < x+h+\Delta} w_{x,h,\Delta}(n)\Lambda(n) = (h + \Delta) \sum_{m^k < h} \frac{1}{m^k} - \frac{1}{\Delta} \sum_{m^k < h} \sum_\rho \frac{S(\rho)}{m^{k\rho}}$$

$$+ O\left(\frac{1}{\Delta x} \sum_{m^k < h} m^{2k}\right). \qquad (4)$$

Estimating by comparison to the integral we have

$$\sum_{m^k < h} \frac{1}{m^k} = \zeta(k) - \sum_{m^k \ge h} \frac{1}{m^k} = \zeta(k) + O(h^{1/k-1}) \qquad (5)$$

8

and

$$\sum_{m^k < h} m^{2k} \ll h^{2+1/k}. \tag{6}$$

We now turn our attention to estimating the sum over the zeroes in (4). Assuming the Riemann hypothesis, we have that

$$\left| \sum_{m^k < h} \sum_{\rho} \frac{S(\rho)}{m^{k\rho}} \right| \ll \left(\sum_{m^k < h} \frac{1}{m^{k/2}} \right) \left| \sum_{\rho} S(\rho) \right|.$$

Clearly, we can write that

$$\sum_{m^k < h} \frac{1}{m^{k/2}} \ll g_k(h)$$

where $g_k(h) = \log h$ for $k = 2$ and $g_k(h) = 1$ for $k \geq 3$. It thus remains to estimate the sum over the zeroes. We split this into three sums by

$$\sum_{\rho} S(\rho) = \left(\sum_{|\gamma| \leq x/h} + \sum_{x/h < |\gamma| \leq x/\Delta} + \sum_{|\gamma| > x/\Delta} \right) S(\rho)$$

and provide bounds in the following lemmas. Standard estimates for sums over the zeroes of $\zeta(s)$ can be found in [2].

Lemma 6. *Assume the Riemann hypothesis and let $2 \leq \Delta \leq h \leq x$. Then*

$$\sum_{|\gamma| > x/\Delta} S(\rho) \ll \Delta x^{1/2} \log x$$

Proof. On the Riemann hypothesis, one has that

$$|S(\rho)| \leq \frac{4(x + h + \Delta)^{3/2}}{\gamma^2}.$$

Therefore we have the bound

$$\sum_{|\gamma| > x/\Delta} S(\rho) \ll x^{3/2} \sum_{\gamma > x/\Delta} \frac{1}{\gamma^2}.$$

The result now follows from the fact that

$$\sum_{\gamma > T} \frac{1}{\gamma^2} \ll \frac{\log T}{T}.$$

$\square$

Lemma 7. *Assume the Riemann hypothesis and let $2 \leq \Delta \leq h \leq x$. Then*

$$\sum_{|\gamma| \leq x/h} S(\rho) \ll x^{1/2} \Delta \log x$$

Proof. We write

$$S(\rho) = \int_{x+h}^{x+h+\Delta} \int_{u-h-\Delta}^{u} t^{\rho-1} dt \, du.$$

Estimating this trivially on the Riemann hypothesis one has

$$
\begin{aligned}
|S(\rho)| &\ll \int_{x+h}^{x+h+\Delta} \int_{u-h-\Delta}^{u} t^{-1/2} dt \, du \\
&\ll \int_{x+h}^{x+h+\Delta} (u-h-\Delta)^{-1/2}(h+\Delta) du \\
&\ll h\Delta x^{-1/2}.
\end{aligned}
$$

Thus,

$$\sum_{|\gamma| \leq x/h} S(\rho) \ll \frac{h\Delta}{x^{1/2}} N(x/h)$$

where $N(T)$ counts the number of zeroes $\rho = 1/2 + i\gamma$ of the Riemann zeta function with $0 < \gamma < T$. The result now follows from the bound

$$N(T) \ll T \log T.$$

$\square$

Lemma 8. *Assume the Riemann hypothesis and let $2 \leq \Delta \leq h \leq x$. Then*

$$\sum_{x/h < |\gamma| \leq x/\Delta} S(\rho) \ll hx^{1/2} \log x$$

Proof. We start by writing

$$S(\rho) = \frac{1}{\rho}\left(\int_{x+h}^{x+h+\Delta} t^{\rho} - (t-\Delta-h)^{\rho} dt\right).$$

We estimate this trivially on the Riemann hypothesis to get

$$S(\rho) \ll \frac{\Delta x^{1/2}}{\gamma}.$$

10

Therefore, we have that

$$\sum_{x/h<|\gamma|\leq x/\Delta} S(\rho) \ll \Delta x^{1/2} \sum_{x/h<|\gamma|\leq x/\Delta} \frac{1}{\gamma}.$$

In consideration of the bound

$$\sum_{\gamma<T} \frac{1}{\gamma} = \frac{1}{4\pi} \log^2 T + O(\log T),$$

it follows that

$$\begin{aligned}
\sum_{x/h<|\gamma|\leq x/\Delta} \frac{1}{\gamma} &= \frac{1}{4\pi}\left(\log^2(x/\Delta) - \log^2(x/h) \right) + O(\log x) \\
&= \frac{1}{2\pi} \int_{x/h}^{x/\Delta} \frac{\log t}{t} dt + O(\log x).
\end{aligned}$$

Estimating the integral by

$$\int_{x/h}^{x/\Delta} \frac{\log t}{t} dt \ll \left(\frac{x}{\Delta} - \frac{x}{h} \right) \frac{\log(x/h)}{x/h} \ll \frac{h}{\Delta} \log x$$

completes the proof of the lemma. $\qquad\square$

We now return to the explicit formula. Combining the estimates from Lemmas 6, 7 and 8 and the bounds (5) and (6) with the explicit formula (4) gives us that

$$\begin{aligned}
\sum_{x-\Delta<nm^k<x+h+\Delta} w_{x,h,\Delta}(n)\Lambda(n) &= \zeta(k)h + O\left(\frac{hg(h)}{\Delta}\sqrt{x}\log x \right) + O(\Delta) \\
&\quad + O(h^{1/k}) + O\left(\frac{h^{2+1/k}}{\Delta x} \right).
\end{aligned}$$

We need to remove, from the above estimate, the contribution that arises from the case where $n = p^r$ and $r \geq 2$. Consider the sum

$$\sum_{\substack{x-\Delta<p^r m^k<x+h+\Delta \\ r\geq 2}} w_{x,h,\Delta}(n) \log p.$$

Clearly, this sum is bounded above by

$$\sum_{\substack{x-\Delta<p^r m^k<x+h+\Delta \\ r\geq 2}} \log p \ll \sum_{m^k<x} \left(\sum_{(x-\Delta)/m^k<p^r<(x+h+\Delta)/m^k} \log p \right).$$

11

It is trivial to bound the inner sum by

$$\sum_{(x-\Delta)/m^k < p^r < (x+h+\Delta)/m^k} \log p \ll \left(\frac{h}{m^{k/2}}\right)^{1/2} \log x,$$

where we have used the fact that the number of r-th powers with $r \geq$ in an interval $(x, x+h)$ is $O(\sqrt{h})$. Thus,

$$\sum_{\substack{x-\Delta < p^r m^k < x+h+\Delta \\ r \geq 2}} w_{x,h,\Delta}(n) \log p \ll h^{1/2} \log^2 x.$$

It follows that

$$\sum_{x-\Delta < pm^k < x+h+\Delta} w_{x,h,\Delta}(n) \log p = \zeta(k)h + O\left(\frac{hg(h)}{\Delta}\sqrt{x}\log x\right) + O(\Delta)$$

$$+ O(h^{1/2} \log^2 x) + O\left(\frac{h^{2+1/k}}{\Delta x}\right).$$

We also need to remove all numbers of the form pm^k that are contained in the intervals $(x - \Delta, x)$ and $(x+h, x+h+\Delta)$. We have by the prime number theorem that

$$\sum_{x-\Delta < pm^k < x} \log p \ll \sum_{m^k < x} \left(\sum_{(x-\Delta)/m^k < p < x/m^k} \log p\right) \ll \sum_{m^k < x} \frac{\Delta}{m^k} \ll \Delta.$$

Estimating the contribution from the interval $(x+h, x+h+\Delta)$ is similar. Therefore, we have that

$$\sum_{x < pm^k < x+h} \log p = \zeta(k)h + O\left(\frac{hg(h)}{\Delta}\sqrt{x}\log x\right) + O(\Delta)$$

$$+ O(h^{1/2} \log^2 x) + O\left(\frac{h^{2+1/k}}{\Delta x}\right).$$

First, let $k \geq 3$ and $h = f(x)x^{1/2} \log x$ where $f(x)$ is a function which goes to infinity as $x \to \infty$. Choosing $\Delta = f(x)^{1/2}x^{1/2} \log x$ gives us that

$$\sum_{x < pm^k < x+h} \log p = \zeta(k)h + O\left(\frac{h}{f(x)^{1/2}}\right). \tag{7}$$

12

In the case where $k = 2$, we let $h = f(x)x^{1/2}\log^2 x$. Then, choosing $\Delta = f(x)^{1/2}x^{1/2}\log^2 x$, we get that

$$\sum_{x < pm^2 < x+h} \log p = \zeta(2)h + O\left(\frac{h}{f(x)^{1/2}}\right). \tag{8}$$

Theorem 3 now follows from (7), (8) and partial summation.

Acknowledgements

References

[1] E. Cohen. Arithmetical notes, IX. On the set of integers representable as a product of a prime and square. *Acta Arith.*, 7(4):417–420, 1962.

[2] H. L. Montgomery and R. C. Vaughan. *Multiplicative number theory I: Classical theory*, volume 97. Cambridge University Press, 2006.